Appeared in **Proc. International Conference on Visual Information**, June 1999.

Automatic Deformable Shape Segmentation for Image Database Search Applications

Lifeng Liu and Stan Sclaroff

Computer Science Dept., Boston University
111 Cummington Street, Boston, MA 02215, USA
liulf@cs.bu.edu, sclaroff@cs.bu.edu

Abstract. A method for shape based image database indexing is described. Deformable shape templates are used to group color image regions into globally consistent configurations. A statistical shape model is used to enforce the prior probabilities on global, parametric deformations for each object class. The segmentation is determined in part by the minimum description length (MDL) principle. Once trained, the system autonomously segments deformed shapes from the background, while not merging them with adjacent objects or shadows. The formulation can be used to group image regions based on any image homogeneity predicate; *e.g.,* texture, color, or motion. Preliminary experiments in color segmentation and shape-based retrieval are reported.

1 Introduction

Retrieval by shape is considered to be one of the more difficult aspects of content-based image database search. A major part of the problem is that many techniques assume that shapes have already been segmented from the background, or that a human operator has encircled the object via an active contour. Such assumptions are unworkable in applications where automatic indexing is required.

In this paper, a new region-based approach is proposed that automatically segments deformable shapes from images. Deformable shape templates are used to group color image regions into globally consistent configurations. A statistical shape model is used to enforce the prior probabilities on global, parametric deformations for each object class. The segmentation is determined in part by the minimum description length (MDL) principle.

The method includes two stages: over-segmentation using a traditional region segmentation algorithm, followed by deformable model-based evaluation of various region grouping hypotheses. During the second stage, region merging, deformable model fitting, and global consistency checking are executed simultaneously. The approach is general, in that it can be used to group image regions based on texture measures, color, or other image features.

Once trained, the system autonomously segments objects from the background, while not merging them with adjacent objects of similar image color. The resulting recovered parametric model descriptions can then be used directly in shape-based search of image databases. The system was tested on a number of different shape classes and results are encouraging.

2 Background

Segmentation using low-level techniques, such as region growing, edge detection, and mathematical morphology operations, requires a considerable amount of interactive guidance in order to get satisfactory results. Automating these model-free approaches is difficult because of shape complexity, illumination, inter-reflection, shadows, and variability within and across individual objects.

One solution strategy is to exploit prior knowledge to sufficiently constrain the segmentation problem. For instance, a model based segmentation scheme can be used to reduce the complexity of region grouping. Due to shape deformation and variation within object classes, a simple rigid model-based approach will break down in general. This realization has led to the use of deformable contour models in image segmentation [11] and in shape-based image retrieval [8, 5].

The snake formulation can be extended to include a term that enforces homogeneous properties over the region during region growing [7, 9, 15]. This region-based approach tends to be more robust with respect to model initialization and noisy data. However, it requires hand-placement of the initial model, or a user-specified seed point on the interior of the region. One proposed solution is to scatter many region seeds at random over the image, followed with segmentation guided via Bayes/MDL criteria [10, 12, 19].

Unfortunately, the above mentioned techniques are going to make mistakes in merging regions, even in constrained contexts. This is because local constraints are in general insufficient. To gain a more reliable segmentation, global consistency must be enforced [17]: the best partitioning is the one that globally and consistently explains the greatest portion of the sensed data. Finding the globally consistent or MDL image labeling is impractical in general due to the computational complexity of global optimization algorithms [13]. This leads to the use of parallel algorithms [12] or algorithms that instead find an approximately optimal solution [2, 4, 6, 10, 14, 16, 18, 19].

3 Model Formulation

In our system, a deformable model is used to guide grouping of image regions. A shape model is specified in terms of global warping functions applied to a closed polygon, hereafter referred to as a template. The global warping can be generic, and is controlled by a vector of warping parameters, $\mathbf{a}$. To demonstrate the approach, we implemented a system that uses quadratic polynomials to model global deformation due to stretching, shearing, bending, and tapering.

Assume that the distribution on shape parameters for a particular shape category can be modeled as a multi-dimensional normal distribution. The distribution is characterized by its mean $\bar{\mathbf{a}}$ and covariance matrix $\mathbf{\Sigma}$. For a given deformation paramter vector $\mathbf{a}$, the sufficient statistic for characterizing likelihood is the Mahalanobis distance:

$$E_{deform} = \tilde{\mathbf{a}}^T \mathbf{\Sigma}^{-1} \tilde{\mathbf{a}}, \tag{1}$$

where $\tilde{\mathbf{a}} = \mathbf{a} - \bar{\mathbf{a}}$. As will be described, $\bar{\mathbf{a}}$, $\mathbf{\Sigma}$ are acquired via supervised learning.

3.1 Model Fitting

One important step in the image partitioning procedure is to fit each region grouping hypothesis $\mathbf{g}_i$ with deformable models from the object library. Fitting minimizes a function that includes the deformation term of Eq. 1 and two additional terms: a.) area overlap between model and region grouping, and b.) color compatibilty of regions included in the grouping:

$$E(\mathbf{g}_i) = E_{color} + \alpha E_{area} + \beta E_{deform}. \tag{2}$$

The scalars α and β control the importance of the three terms. The color compatibility term E_{color} is simply the norm of color covariance matrix for pixels within the region grouping. The region/model area overlap term is computed $E_{area} = \frac{S_G S_m}{S_c^2}$, where S_G is the area of the region grouping hypothesis, S_m is the area of the deformed model, and S_c is the common area between the regions and deformed model. By using degree of overlap in our cost measure, we can avoid measuring distances between region boundaries and corresponding model control points. Hence we can avoid the problem of finding direct correspondence between landmark points, which is not easy in the presence of large deformations.

Model fitting is accomplished by minimizing Eq. 2. In our system, we employ the downhill-simplex method [13] because it requires only function evaluations, not derivatives. Though it is not very efficient in terms of the number of function evaluations that it requires, it is still suitable for our application since it is fully-automatic, and reliable. The procedure is accelerated via a multiscale approach.

3.2 Model Training

In the current system, the template is defined by the operator as a polygonal model. During model training, the system is presented with a collection of color images. These images are first over-segmented via a traditional color region segmentation algorithm [1,13]. In the first few training images, the operator is asked to mark candidate regions that belong to the same object. The system then merges the regions and uses downhill-simplex method to minimize the cost function in Eq. 2, thereby matching the template to the training regions in a particular image. This process is repeated for all images in the training set. As more training data is processed, the system can then semi-automate training. The system can take a "first guess" at the correct region grouping and present it to the operator for approval [13].

4 Automatic Image Segmentation

Once trained, the deformable model guides the grouping and merging of color regions. The process begins with over-segmentation of the color input image [1,13]. An edge map is also computed via standard image processing methods. Using this over-segmentation, candidate regions are matched with models based on their color band-rate feature [3].

There are two major constraints used in the selection of candidate groupings. The first constraint is a spatial constraint: every region in a grouping hypothesis should be adjacent to another region in the same group. The second constraint is a region boundary compatibility constraint [13]: if the boundary between two region is "strong," then they cannot be combined in the same group.

The system then tests various combinations of candidate region groupings for each model. The goal is to find the optimal, model-based partitioning of the image. In theory, the system should exhaustively test all possible combinations of the candidate regions, and select the best ones for merging; however, the computational complexity of such exhaustive testing is exponential, and the problem of finding the best group is NP complete. To make the problem tractable, we have tested a number of approximation strategies for finding an globally consistent labeling of the image [13].

In the global consistency strategy, for any possible partitioning of the image, we compute a global cost value for the whole configuration:

$$\mathcal{E} = \sum_{i=1}^{n} \mathbf{r}_i \mathbf{E}(\mathbf{g}_i) + \gamma \mathbf{n}, \tag{3}$$

where $\mathbf{r}_i$ is the ratio of $\mathbf{i}^{th}$ group area to the total area, and $\mathbf{E}(\mathbf{g}_i)$ is the deformation cost for group $\mathbf{g}_i$, $\mathbf{n}$ is the number of the groupings in the current image partitioning, and γ is a constant factor. In our experiments, $\gamma = 0.04$.

The first term measures the model compatibilty over all groupings in the image partition. The second term corresponds to the code length (number of models employed); it enforces a minimum description length criterion [12,13].

4.1 Highest Confidence First

A deterministic algorithm, highest confidence first (HCF), can be used to improve convergence speed [4,10]. The HCF algorithm as applied to our problem is as follows:

1. Initialize the region grouping configuration such that every region in the over-segmented image is in its own distinct group $\mathbf{g}_i$.
2. Fit models to each region grouping $\mathbf{g}_i$. Compute the global cost $\mathcal{E}_o$ via Eq. 3. Save this configuration as best found so far, C_o.
3. Set $\mathcal{E}_m$ to a very large value.
4. For each pair of adjacent groups g_i, g_j in the current configuration, compute the global cost, $\mathcal{E}_2$ that would result if g_i, g_j were merged. If $\mathcal{E}_2 < \mathcal{E}_m$, then set $\mathcal{E}_m = \mathcal{E}_2$ and save this merged configuration C_m. After this step, C_m is the configuration with minimum merging cost for merging any pair of groups in the current configuration.
5. Use the merged configuration C_m as the new configuration. If $\mathcal{E}_m < \mathcal{E}_o$, then set $\mathcal{E}_o = \mathcal{E}_m$ and save this new configuration as best found so far $C_o = C_m$.
6. Terminate when all groups are merged into one. and output the best configuration C_o and its cost value $\mathcal{E}_o$. Otherwise, go to 3.

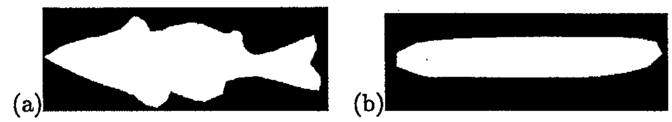

Fig. 1. Two deformable template models employed in our experiments: (a) fish model, (b) banana model. The initial polygonal model was defined by the user, and then trained as described in Sec. 3.2.

In our experience, the computational complexity of HCF is generally less than that needed to obtain similar quality segmentation results via the simulated annealing algorithm [13]. In HCF the number of different merging configurations tested is $\mathbf{O}(\mathbf{n}^2)$, where $\mathbf{n}$ is the number of regions in the image. This is because some results from the previous iteration can be reused in the next. Specifically, at each iteration (except the first), the algorithm need only compute the pairwise merging cost between all groups g_i and the newly-merged group from the previous iteration.

5 Results

The aforementioned segmentation method was implemented and tested on hundreds of images from a number of different classes of cluttered color imagery: images of fruit, vegetables, and leaves collected under controlled lab conditions, and images of fish obtained from the world wide web. Due to space limitations, only two examples can be shown.

The first example shows segmentation results for five examples of fish images obtained from the world wide web. The fish model used in segmentation is shown in Fig. 1(a), and was trained using about 60 training images. The test images were excluded from the training set. The original color images are shown in the first column of Fig. 2, followed by the over-segmented images used as input to the merging algorithm. The third column shows the models recovered in finding the best merging configuration obtained via HCF. Finally, last column depicts the corresponding model-based merging of image regions.

As can be seen, the method accurately recovered a deformable model description of each fish in the image. Only in one case, (Fig. 2(a)), was the orientation of some of the models incorrectly estimated. Despite clutter, deformation, and partial occlusions, performance was quite satisfactory.

In the next example, we show the approach as employed in an image retrieval application. We demonstrate the approach using a simple banana shape model that was trained using 40 example images of bananas at varying orientations and scales. These training images were not contained in our test image data set.

All images in the test data set were then segmented using the trained model as described in Sec. 4. The recovered model deformation parameters $\mathbf{a}$ for the selected region grouping hypotheses were stored in the index for each image. If the image had multiple yellow objects, then the system stored a list of model

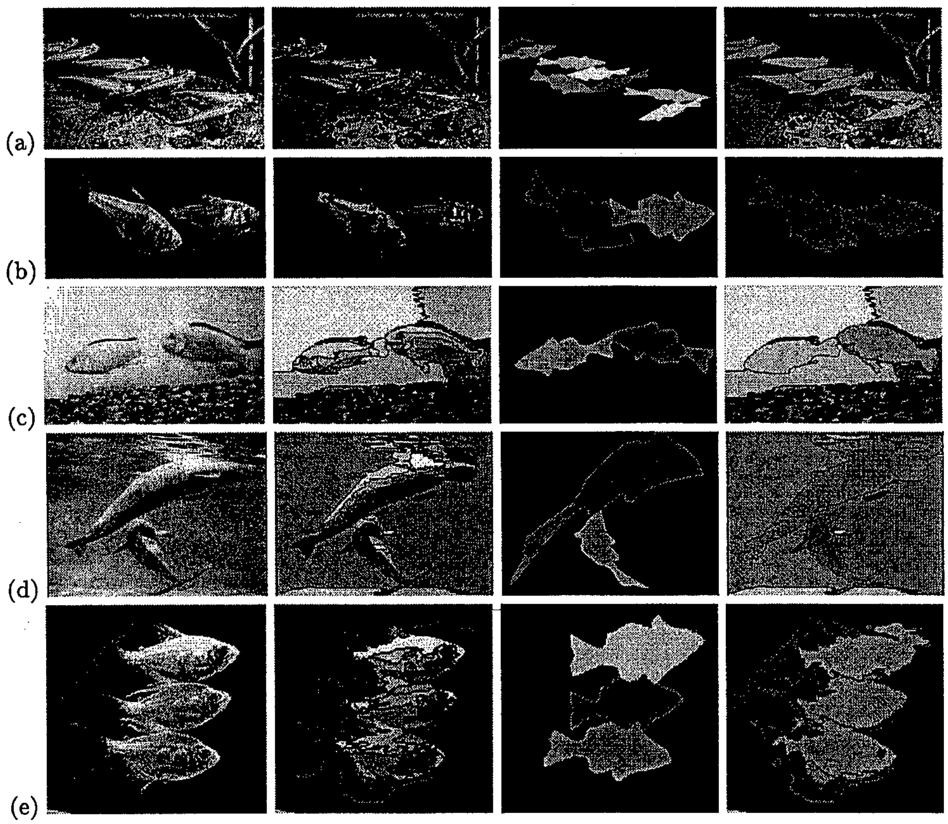

Fig. 2. Example segmentation for images of fish. The original color images are shown in the first column, followed by the over-segmented images used as input to the merging algorithm. The third column shows the recovered deformable models for the best merging configuration obtained via HCF. Finally, last column depicts the model-based merging of regions.

descriptions for that image. Once descriptions are precomputed, shape-based queries can be answered in interactive time.

An example search with our system is shown in Fig. 3. The user selected the image shown in Fig. 3(0). The system retrieved images that had similar shapes, here shown in rank order (1-14). The most similar shapes are other bent bananas of similar aspect ratio. Yellow squash shapes were ranked less similar. The corresponding region grouping is shown below each of the original images in the figure.

Note that the system correctly grouped regions despite shadows, lighting conditions, and deformation. Especially notable are cases where multiple yellow shapes are abutting each other (Fig. 3(3,7,12,14)). Due to the use of model-based region merging, our system is able to avoid merging similarly colored, adjacent but separate objects. The approach is also adept at avoiding merging objects with their similarly-colored shadows.

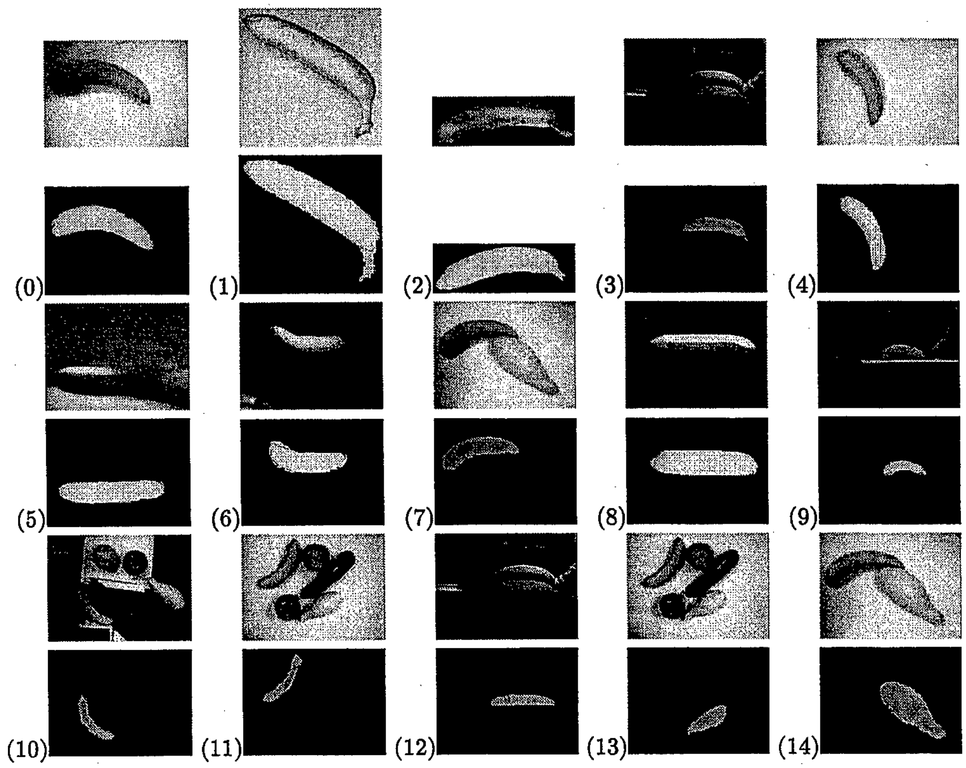

Fig. 3. Image retrieval example. The user selected an example image (0). The system retrieved shapes found in the database and displayed them in rank similarity order (1-14). The segmented shape is shown below each original database image. If an image contained more than one yellow shape, it is shown more than once in the retrieval (once per shape). Note that the most similar shapes are other bent bananas of similar aspect ratio. Yellow squash shapes were ranked less similar.

6 Conclusion

As seen in the examples of the previous section, the shape-based region merging algorithm can produce satisfactory results. The algorithm can detect the whole object correctly, while at the same time, avoid merging objects with background and/or shadows, or merging adjacent multiple objects. A statistical shape model is used in finding a globally-consistent labeling of the image, as determined in part by the minimum description length (MDL) principle. The formulation is general, in that it can be used to group image regions based on a general image homogeneity predicate; *e.g.,* texture, color, or motion.

The major issue is the computation time required to obtain a segmentation result. This led to the evaluation of different methods for obtaining approximate, globally optimal region groupings [13]. The method of choice is based upon the highest confidence first (HCF) algorithm.

In most previous approaches, initial model placement is either given by the operator, or by exhaustively testing the model in all orientations, scales, and deformations centered at every pixel in the image. The region-based approach proposed in this paper significantly reduces the need to test all model positions. Once trained, our system is fully-automatic. Therefore, it is well-suited to image database indexing applications. Each selected region grouping hypothesis has a recovered shape model associated with it. As has been demonstrated, these model parameters can be used directly in recognition and shape comparison.

References

1. J.R. Beveridge, J.S. Griffith, R.R. Kohler, A.R. Hanson, and E.M. Riseman. Segmenting images using localized histograms and region merging. *IJCV*, 2(3):311–352, 1989.
2. G. Bongiovanni and P. Crescenzi. Parallel simulated annealing for shape detection. *CVIU*, 61(1):60–69, 1995.
3. M. H. Brill. Can color-space transformation improve color constancy other than von Kries? *SPIE, Human, vision, visual Processing, and digital display TV*, 1913:485–492, 1993.
4. P. B. Chou and C. M. Brown. The theory and practice of bayesian image labeling. *IJCV*, 4(3):185–210, 1990.
5. A. DelBimbo and P. Pala. Visual image retrieval by elastic matching of user sketches. *PAMI*, 19(2):121–132, 1997.
6. R. P. Grzeszczuk and D. N. Levin. Brownian strings: segmentating images with stochastically deformable contours. *PAMI*, 19(10):1100–1114, 1997.
7. J. Ivins and J. Porrill. Active-region models for segmenting textures and colors. *I&VC*, 13(5):431–438, 1995.
8. A. K. Jain, Y. Zhong, and S. Lakshmanan. Object matching using deformable templates. *PAMI*, 18(3):267–278, 1996.
9. T. N. Jones and D. N. Metaxas. Image Segmentation Based on the Integration of Pixel Affinity and Deformable Models. *Proc. CVPR*, pp. 330–337, 1998.
10. T. Kanungo, B. Dom, W. Niblack, and D. Steele. A fast algorithm for mdl-based multi-band image segmentation. *Proc. CVPR*, pp. 609–616, 1994.
11. M. Kass, A.P. Witkin, and D. Terzopoulos. Snakes: Active contour models. *IJCV*, 1(4):321–331, 1988.
12. Y. G. Leclerc. Constructing simple and stable descriptions for image partitioning. *IJCV*, 3(1):73–102, 1989.
13. L. Liu and S. Sclaroff. Deformable shape detection and description via model-based region grouping. Technical report, CS TR 98-017, Boston U., Nov. 1998.
14. D. Noll and W. Von Seelen. Object recognition by deterministic annealing. *I&VC*, 15(11):855–860, 1997.
15. R. Ronfard. Region-based strategies for active contour models. *IJCV*, 13(2):229–251, 1994.
16. G. Storvik. Bayesian approach to dynamic contours through stochastic sampling and simulated annealing. *PAMI*, 16(10):976–986, 1994.
17. T. M. Strat. *Natural Object Recognition*. Springer-Verlag, 1992.
18. J. P. Wang. Stochastic relaxation on partitions with connected components and its application to image segmentation. *PAMI*, 20(6):619–636, 1998.
19. S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation. *PAMI*, 18(9):884–900, 1996.

REPORT DOCUMENTATION PAGE

OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
August 1999

3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE
Automatic Deformable Shape Segmentation for Image Database Search Applications

5. FUNDING NUMBERS
G N00014-96-1-0661

6. AUTHOR(S)
Lifeng Liu and Stan Sclaroff
Stan Sclaroff

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Computer Science Department
Boston University
111 Cummington Street
Boston, MA 02215

8. PERFORMING ORGANIZATION REPORT NUMBER
sclaroff-ONR-TR99-777

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Department of the Navy
Office of Naval Research
Ballston Centre Tower One
800 North Quincy Street
Arlington, VA 22217-5660

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)
A method for shape based image database indexing is described. Deformable shape templates are used to group color image regions into globally consistent configurations. A statistical shape model is used to enforce the prior probabilities on global, parametric deformations for each object class. The segmentation is determined in part by the minimum description length (MDL) principle. Once trained, the system autonomously segments deformed shapes from the background, while not merging them with adjacent objects or shadows. The formulation can be used to group image regions based on any image homogeneity predicate; e.g., texture, color, or motion. Preliminary experiments in color segmentation and shape-based retrieval are reported.

14. SUBJECT TERMS
Non-rigid shape description and recognition; image and video database modeling, content-based retrieval

15. NUMBER OF PAGES
8

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT
unclassified

18. SECURITY CLASSIFICATION OF THIS PAGE
unclassified

19. SECURITY CLASSIFICATION OF ABSTRACT
unclassified

20. LIMITATION OF ABSTRACT
UL